**PCT** WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12Q 1/68

A1

(11) International Publication Number: **WO 00/53811**

(43) International Publication Date: 14 September 2000 (14.09.00)

(21) International Application Number: PCT/US00/06342

(22) International Filing Date: 9 March 2000 (09.03.00)

(30) Priority Data:

| | | |
|---|---|---|
| 60/123,878 | 11 March 1999 (11.03.99) | US |
| 09/521,391 | 8 March 2000 (08.03.00) | US |

(71) Applicant: ORION GENOMICS, LLC [US/US]; 4041 Forest Park Blvd., St. Louis, MO 63108 (US).

(72) Inventors: MCPHERSON, John, D.; 9017 State Route 162, Troy, IL 62294–2070 (US). WILSON, Richard, K.; 2234 Whitney Pointe Drive, Chesterfield, MO 63005–4515 (US). MENG, Xun; Apt. F, 12362 Amber Creek Court, St. Louis, MO 63141–6430 (US).

(74) Agents: MCDANIEL, C., Steven et al.; McDaniel & Associates P.C., P.O. Box 2244, Austin, TX 78768–2244 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

**Published**
*With international search report.*

(54) Title: GENOME CHIPS AND OPTICAL TRANSCRIPT MAPPING

(57) Abstract

A genomic DNA microarray chip is constructed with an ordered, tiled array of oligonucleotide genomic DNA clones having minimally overlapping sequences in a sequence that mimics the selected genome. These genomic DNA microarray chips are used as a tool for visual transcription profiling and visual gene mapping.

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **AL** | Albania | **ES** | Spain | **LS** | Lesotho | **SI** | Slovenia |
| **AM** | Armenia | **FI** | Finland | **LT** | Lithuania | **SK** | Slovakia |
| **AT** | Austria | **FR** | France | **LU** | Luxembourg | **SN** | Senegal |
| **AU** | Australia | **GA** | Gabon | **LV** | Latvia | **SZ** | Swaziland |
| **AZ** | Azerbaijan | **GB** | United Kingdom | **MC** | Monaco | **TD** | Chad |
| **BA** | Bosnia and Herzegovina | **GE** | Georgia | **MD** | Republic of Moldova | **TG** | Togo |
| **BB** | Barbados | **GH** | Ghana | **MG** | Madagascar | **TJ** | Tajikistan |
| **BE** | Belgium | **GN** | Guinea | **MK** | The former Yugoslav Republic of Macedonia | **TM** | Turkmenistan |
| **BF** | Burkina Paso | **GR** | Greece | | | **TR** | Turkey |
| **BG** | Bulgaria | **HU** | Hungary | **ML** | Mali | **TT** | Trinidad and Tobago |
| **BJ** | Benin | **IE** | Ireland | **MN** | Mongolia | **UA** | Ukraine |
| **BR** | Brazil | **IL** | Israel | **MR** | Mauritania | **UG** | Uganda |
| **BY** | Belarus | **IS** | Iceland | **MW** | Malawi | **US** | United States of America |
| **CA** | Canada | **IT** | Italy | **MX** | Mexico | **UZ** | Uzbekistan |
| **CF** | Central African Republic | **JP** | Japan | **NE** | Niger | **VN** | Viet Nam |
| **CG** | Congo | **KE** | Kenya | **NL** | Netherlands | **YU** | Yugoslavia |
| **CH** | Switzerland | **KG** | Kyrgyzstan | **NO** | Norway | **ZW** | Zimbabwe |
| **CI** | Côte d'Ivoire | **KP** | Democratic People's Republic of Korea | **NZ** | New Zealand | | |
| **CM** | Cameroon | | | **PL** | Poland | | |
| **CN** | China | **KR** | Republic of Korea | **PT** | Portugal | | |
| **CU** | Cuba | **KZ** | Kazakstan | **RO** | Romania | | |
| **CZ** | Czech Republic | **LC** | Saint Lucia | **RU** | Russian Federation | | |
| **DE** | Germany | **LI** | Liechtenstein | **SD** | Sudan | | |
| **DK** | Denmark | **LK** | Sri Lanka | **SE** | Sweden | | |
| **EE** | Estonia | **LR** | Liberia | **SG** | Singapore | | |

# GENOME CHIPS AND OPTICAL TRANSCRIPT MAPPING

## FIELD OF THE INVENTION

The present invention relates to genomic DNA microarray chips constructed with an ordered tiled array of oligonucleotide genomic DNA clones having minimally overlapping sequences. More specifically, the present invention relates to the use of the genomic DNA microarray chips as a tool for transcription profiling and gene discovery.

## BACKGROUND OF THE DISCLOSURE

Essential to the analysis of the genetic complement of an organism and to the exploitation and manipulation of its genome is the identification of the population of genes expressed under varying circumstances. The elucidation of the up- or down-regulation of any particular gene or genes when the organism is challenged provides putative commercial gene targets. These changes may be due to, or the cause of, a disease state in the organism, or a response to chemical or physical stimuli.

Several methods for analyzing gene expression exist; however, only hybridization-based approaches are amenable to the multiplex analysis of complete sets of RNA transcripts derived from an organism or tissue. Nucleic acid hybridization technology has evolved from Southern's initial observation that complementary base-pairing could be exploited for the interrogation of nucleic acid molecules immobilized on a solid support by using nucleic acid molecules labeled with a reporter molecule as a probe (Southern, 1975).
The technique was extended to the screening of a collection of DNA clones, called a clone library, replicated onto nitrocellulose or nylon filters thereby allowing a direct correlation between clones and observed signals from the hybridizing probes. Individual clones from a clone library were eventually stored in unique wells of a series of microtiter plates. These arrayed libraries have become the standard for positional cloning work as each clone can be reproducibly replicated onto a filter in a fixed position allowing data associated with each clone to be accumulated from multiple experiments. If the clones in the arrayed library represent genes, then probes derived from RNA transcripts can be used to explore gene expression (Schena, 1995; Duggan, 1999, Brown, 1999).

To date, the most successful methodologies utilize DNA immobilized on nonporous supports such as glass. This has facilitated the use of fluorescent reporter molecules and

small-volume hybridizations, thus allowing miniaturization. Methods for robotically placing DNAs (Schena, 1995) or directly synthesizing oligonucleotides (Fodor, 1991) in spatially defined high-density arrays have been devised to produce what are often referred to as DNA microarrays or gene chips. Three typical sources for the immobilized DNA used for these chips include oligonucleotides designed from known gene sequences, inserts from cDNA clones, and PCR amplified products derived from computationally-predicted coding sequences identified in genomic sequence data. All three of these sources require prior gene discovery as they utilize fragments of identified coding sequences. As such, the sets of genes represented on the currently utilized chips are likely to be incomplete for all but perhaps totally sequenced genomes (e.g., yeast) and further will require knowledge of all coding sequences before they can be fully represented. The size and complexity of most genomes of commercial interest make the latter approach extremely expensive and time-consuming.

Mapping and DNA sequence analysis is currently underway for the human genome, as well as for genomes of important model organisms such as mouse, the roundworm *C. elegans*, and the fruitfly *D. melanogaster*. As these data become available, their usefulness in assaying gene expression will be largely dependent on identifying all of the genes contained within using computational tools and traditional biological and biochemical methods. This will be time-consuming, expensive and, for many genes, inaccurate and unsuccessful.

There is an existing need for a method of assaying gene expression which can include the complete set of genes from an organism that will be much more effective in pinpointing all of the genes responsible for a specific disease state or in the response to a specific environmental stimulus. The conventional DNA microarrays or gene chips currently being developed and used for this purpose will not be effective as they rely on prior characterization of every gene-coding segment of a genome.

## SUMMARY OF THE INVENTION

The disclosure teaches a method to prepare genomic DNA microarray chips that contain arrays of ordered genomic DNA clones that serve as tools for gene expression analysis. One embodiment of the disclosed invention is a genomic DNA microarray chip, constructed with an array of oligonucleotide sequences, which is immobilized on a solid support. This array is comprised of a group of genomic DNA clones having minimally

overlapping sequences and together make up an ordered, tiling path representing at least part of an entire genome with at least one of the oligonucleotide sequences being of an unknown sequence.

A second embodiment is a method of determining the genomic location of a labeled known DNA sequence. This is accomplished by hybridizing a known labeled DNA sequence chosen from the group consisting of an EST, cDNA, PCR product or genomic fragment to at least one DNA clone on the immobilized genomic DNA microarray chip.

A third embodiment of the disclosed invention is a method of using the genomic DNA microarray chips for initial gene discovery and characterization or visual transcript mapping.

A fourth embodiment is a method of preparing the array of DNA clones that will be placed on the genomic DNA microarray chip. This comprises the isolation and fragmentation of the DNA, the cloning of each of these genomic DNA fragments, isolating the individual clones of the genomic DNA fragments, identifying their ordered tiling path and immobilizing the DNA fragments in a sequential alignment representing their original position in the genome

Another embodiment is the construction of a genomic DNA microarray chip using an immobilized array of DNA clones, which represent oligonucleotide, sequences of substantially the entire genome. The DNA clones are arranged in an order reflecting the original sequence of oligonucleotides of the genome

Yet another embodiment is a kit for determining the genomic location of known oligonucleotide sequences that comprises an immobilized ordered array of DNA clones derived from a plurality of genomic DNA fragments and reflects the original sequence of the genome from which it was isolated. The kit would also include a labeled sample of a known oligonucleotide sequence wherein said known oligonucleotide sequence will hybridize to at least one of the immobilized genomic DNA clones on the microarray chip.

Another embodiment of the disclosure is a genomic DNA clone array comprised of a multiwelled plate having one DNA clone in each well of this plate. Each genomic DNA clone was selected from a plurality of cloned genomic DNA fragments and has an overlapping nucleotide sequence in common with the genomic DNA clone in at least one adjacent well.

Yet another embodiment of the disclosure is the use of the genomic DNA microarray chip as a tool for the analysis of gene expression, visual transcriptional profiling and visual transcription mapping.

The foregoing has outlined rather broadly the features and advantages of the present invention in order that the detailed description of the invention that follows may be better understood. Additional features and advantages of the invention will be described hereinafter which form the subject of the claims of the invention.

## BRIEF DESCRIPTION OF THE DRAWINGS

The novel features which are believed to be characteristic of the invention will be better understood from the following detailed description, in conjunction with the accompanying drawings.

**Figure 1.** Illustrates genomic DNA microarray chips used for transcription profiling.

**Figure 2A.** Illustrates a genome containing multiple chromosomes.

**Figure 2B.** Illustrates a chromosome and a minimal tiling path for a selected region of that chromosome.

**Figure 3.** Illustrates the alignment of DNA clones on a DNA microarray chip.

**Figure 4.** Illustrates the identification of known genes in a genomic DNA sequence using predicted exons, protein similarities and EST matches.

**Figure 5.** Illustrates the results of hybridized target molecules to minimally overlapping DNA clones.

**Figure 6.** Illustrates visualized genes on individual genomic clones.

**Figure 7.** Illustrates the process of visual transcription profiling.

## DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

As illustrated in Figure 1 the present invention is directed to the construction and use of genomic DNA microarray chips 10 where the genomic DNA clones are adhered to discrete areas 12 on a solid support surface. The genomic DNA clones are adhered in an ordered array of overlapping, tiled genomic DNA cloned sequences starting at the upper left hand corner of the micro chip and proceeding left to right to the bottom of the chip. The order of the array mimics the order of the original genomic sequence.

There is some ambiguity in the scientific literature as to the relevant nomenclature, so it is important to define some specific terms within this disclosure. In the context of chip-based DNA array technologies, the term probe is typically used to describe DNA immobilized onto a solid support, while target is used to describe the labeled molecules that are being queried. A DNA clone is a recombinant DNA molecule that has been replicated autonomously in a suitable host cell. The term oligonucleotide sequence refers to a fragment of DNA sequence of any length. An oligonucleotide array is defined as an ordered progression of oligonucleotide sequences. A tiled oligonucleotide array is one where the oligonucleotide sequences represent all or some the oligonucleotides of the original selected genomic DNA and are arranged in a manner that represents the exact order of the selected genomic DNA. A contig is a set of overlapping contiguous clones that cover a chromosome region or a whole chromosome. In order to represent all the oligonucleotides in tiled arrays, small areas of the 5' and 3' ends of the arrayed oligonucleotide sequences will be duplicated or overlapped within the tiled arrayed clones. A microarray chip is defined as a miniaturized oligonucleotide array usually adhered to a solid surface such as glass. These designations will be followed throughout this disclosure except where specifically noted to the contrary.

The following examples are given to provide an increased understanding of the invention as examples, not as limitations to the invention. For example, a microarray chip may be constructed to contain the entire genomic DNA or any portion thereof, even though the example describes the process for the entire genome.

EXAMPLE 1

1. Preparation of Genomic DNA Oligonucleotide Arrays

To provide the basis for the genomic DNA microarray chips, physical genomic DNA oligonucleotide arrays were constructed for the genome of interest using large-insert clones and employing the rapid restriction fingerprinting techniques developed by the present inventors (Marra et al., 1997). The bacterial artificial chromosome (BAC) cloning system developed by Shizuya et al. (*PNAS* 89, 8794-8797, 1992) was shown to accept and maintain stable human and plant DNA fragments up to a size of 350 kb. In contrast to YACs, a yeast-based vector system for cloning large DNA inserts, this F-factor-derived vector is propagated in *E.coli*. BACs show several favorable characteristics such as a low frequency of chimeric clones, easy handling of clones and libraries (e.g. propagation, plating, storage, or colony

hybridization), and simple purification of the cloned DNA. While YACs due to their large insert sizes (up to more than 1 Mb) are still indispensable for the generation of physical maps of very large (> 500 Mbp) genomes, BACs will serve as preferred resources for map based cloning and large scale genome sequencing. BAC clones are preferred for the physical mapping, in this disclosure although other genomic clones could be used as well with some modifications to the workflow. A BAC library for the probe genome should contain at least 15 genome equivalents.

Figure 2A illustrates a target genome 20 that is comprised of multiple chromosomes 24. DNA fragments are generated, cloned and ordered into an ordered genomic array that can represent the coding sequences for all or part of the target genome. DNA is purified from each clone in the BAC library and fingerprinted by restriction digestion. Following agarose gel electrophoresis, the restriction digests are imaged and the fingerprint data used to computationally determine the relationships between all of the clones. The resulting relationship matrix represents the clone-based physical map of the original genome. From this matrix, a minimal set of overlapping, tiled clones, representing the original genome laid end to end can be selected.

Figure 2B shows a map consisting of ordered DNA genomic clones from a selected region 26 of a chromosome 24 of a target genome 20. Figure 2B illustrates the ordered array of the entire genomic DNA cloned sequences 22 found in the designated fraction 26 of the chromosome 24. Only a small portion of the total genomic DNA cloned sequences are selected for the final ordered array.

The chosen DNA cloned sequences 28 (all the bold lines) mimic the final ordered array of the original genome. These chosen DNA cloned sequences 28 are contigs and are shown in Figure 2B as an ordered overlapping tiled array. A tiled array is made of contigs, each contig containing small regions of overlapping and identical sequences to the contig on its left at its 5' left end and to the contig on its right at its 3' right end. These overlapping sequences for contig 29 are indicated with dotted lines in Figure 2B

Briefly, high molecular weight DNA will be isolated from a genome of interest, randomly sheared or partially cleaved using restriction enzymes and ligated into a suitable vector. After transfer of the ligation products into a suitable bacterial host, individual clones will be isolated and arrayed in microtiter plates to construct a genomic library. The combined inserts of this library will encompass approximately 15-fold DNA coverage of the

genome. Each clone from this library will then be grown in culture and its DNA extracted. An aliquot of each DNA will then be digested to completion with a restriction enzyme such as Hind III and the resulting fragments separated according to size by agarose gel electrophoresis. An image of the gel will then be captured using a scanning fluorescence detector and the size of all fragments in each lane determined relative to fragments of known size from a marker that is co-electrophoresed with the digested DNA samples. Groups, or contigs of ordered, overlapping clones are then assembled by comparing restriction pattern similarities or fingerprints, assuming that clones sharing a majority of similarly sized fragments originate from the same portion of the genome. A minimally overlapping set of clones can be selected as an ordered tiling path representing the genome. For the human genome, which consists of over 3 billion subunits or base pairs, a tiling path would be comprised of approximately 340,000 BAC clones.

A. Preparation of DNA

The following is one method of isolation and purification of a recombinant DNA vector from the host organism in which it was placed for amplification.

Culture volumes of 1200 µl of 2X YT (Sambrook et al. 1989) containing 12.5 µg/ml of chloramphenicol (Sigma; fosmids and bacterial artificial chromosomes (BACs)) or kanamycin (Sigma; P1-derived artificial chromosomes (PACs) clones) or the appropriate quantity of antibiotic for the desired clones are inoculated with a single colony which contains one unique recombinant DNA vector from a freshly streaked plate. If desired, multiple single isolate colonies can be processed individually for comparison. Cultures are grown in 2-ml 96-well blocks (Beckman; part 140504) for 24 hr at 37°C with agitation at 300 rpm in a Labline incubator shaker. After growth, glycerol stocks in 96-well format are prepared by combining 50 µl of 80% glycerol with 100 µl of culture and mixing with a 12-channel pipettor. The microplates are sealed with Scotch-brand heavy-duty aluminum foil tape and stored at -80°C.

Bacterial cell cultures (96-well, 1.9 ml from above) are pelleted by centrifugation at 2700 rpm for 15 min in a Jouan model GR-422 floor centrifuge fitted with microplate carriers. The supernatant is decanted away from the pellet, and the 96-well block inverted on a paper towel for 5 min to drain excess culture media. The inverted block is rapped vigorously on a fresh paper towel until excess culture media is removed and then placed

immediately on ice. Alternatively, after removal of the culture media, blocks are sealed with foil tape and stored at -80°C until DNA preparation can be performed.

DNA preparation is performed using a modified alkaline lysis procedure (Sambrook et al. 1989). The cell pellet is resuspended by addition of 50 µl of chilled GET/RNase buffer [50 mM glucose, 25 mM Tris-HCl (pH 8.0), 10 mM EDTA (pH 8.0), 0.12 mg/ml RNase (Sigma R6513)] and vigorous vortexing. After the pellet is thoroughly resuspended an additional 150 µl of GET/RNase is added followed by gentle vortexing to mix. Cell lysis is achieved by addition of 200 µl of a mixture containing 0.2 N NaOH/1% SDS (freshly prepared), rotation of the block 90° along its long axis 20 times, followed by incubation on the bench for 5 min. Ice cold 3 M potassium acetate (200 µl) (KAc; pH 5.5) is then added to each well, the block tightly sealed with foil tape, and rapidly inverted three times before a 10-min incubation in ice water. For fosmid clones, cleaner DNA preparations, as assayed by examination of digested DNA run on agarose gels, are achieved using 3 M KAc (pH 4.9). However, use of this reagent for purification of BAC DNA invariably results in reduced yield compared to KAc at higher pH. The taped block is inverted rapidly once after the 10-min incubation. Cell debris is then pelleted by centrifugation of the block for 15-20 min at 4000 rpm in a Jouan GR-422 centrifuge maintained at a temperature of 4°C. After centrifugation, blocks are immediately placed on ice. During the last few minutes of the centrifugation, 600 µl of isopropanol are added to each well of a fresh 96-well block (Beckman). This isopropanol-filled block is then inserted into a vacuum manifold (Qiavac 96; Qiagen) and a Qiafilter 96 filter (Qiagen, part 19663) is placed on top of the manifold in preparation for filtration of the supernatant-containing DNA.

After centrifugation, supernatant-containing DNA is separated from the cell debris by inserting a 12-channel pipettor into the block until the tips touch the bottoms of the well. Moving the tips slightly creates a channel in the cell debris, which facilitates removal of the supernatant while leaving the majority of the debris in the well. The supernatant is then transferred to a Qiafilter. When transfer of the supernatant is complete, a vacuum is applied to the Qiafilter manifold, which serves to draw the supernatant through the Qiafilter into the isopropanol-containing block positioned below. In this way residual SDS/cellular debris, which was not pelleted during centrifugation, is removed.

The block is then tightly sealed with foil tape and inverted rapidly three times to mix the supernatant and isopropanol. Precipitation of the DNA is achieved by room temperature

incubation for 15 min followed by a 30-min centrifugation at 4000 rpm. The foil tape is removed and the block inverted to remove the supernatant. The DNA pellet is then washed with 200 µl of 80% ethanol added to the side of the well, and then collected in the bottom of the well by a 10-min centrifugation at 4000 rpm after sealing the block with foil tape. The tape is removed and the block inverted on paper towels for 5 min to drain excess ethanol away from the pellet. The block is then placed in a Savant DNA 110 SpeedVac set at medium heat for 5 min to dry the DNA. The dried pellet is resuspended in 30 µl of TE [10 mM Tris-HCl (pH 8.0), 0.1 mM EDTA (pH 8.0)] in the case of fosmid, BAC, PAC P1 clones, or 150 µl of TE for cosmid clones. Resuspension of the DNA is achieved by incubating the sealed block for 30 min in a 37°C water bath followed by brief vortexing. The DNA is collected in the bottom of the wells by a brief centrifugation and transferred to a nontissue culture-treated microplate that is sealed with foil tape for storage at 20°C.

Alternatively, DNA is prepared by serial addition of 150 µl each of GET/RNase, SDS/NaOH, and KAc pH 5.5 as described above. After addition of KAc, the sealed block is inverted gently three times and then placed in ice water for at least 10 min. The block is inverted twice vigorously before centrifugation, as described. While samples are undergoing centrifugation, 330 µl of 100% ethanol is aliquoted into each well of a 96-well polystyrene Uni-Filter 800 receiver plate (Polyfiltronics). A 0.45 µM cellulose acetate 96-well filter plate (Polyfiltronics) is then mounted on top of the receiver plate and taped securely in place. After centrifugation, a 12-channel pipette (Costar) is used to transfer 400 µl of supernatant-containing DNA to the 96-well filter plate mounted on top of the receiver plate. The assembly, consisting of filter plate and receiver plate, is then subjected to an additional centrifugation at 4000 rpm for 15 min. After centrifugation, the filter plate assembly is dismantled and the ethanol decanted. The DNA pellet is washed with 250 µl of 80% ethanol, dried, and resuspended in the appropriate volume of 10 mM Tris-HCl, 0.1 mM EDTA. This alternative procedure has the advantage of being somewhat more rapid and substantially less expensive due to the use of Polyfiltronics plasticware.

B. Restriction Enzyme Digestion

Restriction endonucleases are used to cut the genomic DNA sequences, which were inserted into the vector and amplified, into smaller fragments based on the location of the specified restriction sites of the chosen enzyme.

For PAC, P1 and BAC DNAs, individual restriction digests should consist of 3.75 µl of ddH20, 1 µl of 10X buffer B (Boehringer-Mannheim), 0.25 µl of HindIII (40 U/µl; Boehringer-Mannheim), and 5 µl of DNA. For fosmid clones, individual restriction digestions should contain 2.75 µl of ddH20, 1 µl of 10X Buffer B (Boehringer-Mannheim), 0.125 µl of HindIII (40 U/µl; Boehringer-Mannheim), 0.1 µl of PstI (100 U/µl; NEB), and 6 µl of DNA. For cosmid clones, each digest should contain 6.75 µl of ddH20, 1 µl 10X buffer B, 0.25 µl of HindIII (40 U/µl; Boehringer-Mannheim), and 2 µl of DNA. The DNA prepared as described above is not quantitated. Yields are usually uniform and the volumes indicated for digestion are adequate. Components of the digestion cocktail are assembled in 96-well thin wall cycle plates (Robbins Scientific). Digestion is achieved by incubation of the cycle plates at 37°C for 4.5 hr in a 96-well thermocycler (MJ Research). After digestion a brief centrifugation collects the DNA in the bottom of the wells and 2 µl of 6x loading dye (0.25% bromophenol blue, 0.25% xylene cyanol FF., 15% Ficoll; Sambrook et al. 1989) is added to each well. Cycle plates are sealed with foil tape and stored at 4°C overnight before agarose gel electrophoresis.

C. Agarose Gel Electrophoresis and Data Acquisition

The restricted genomic DNA fragments were separated by agarose gel electrophoresis stained, scanned and visualized for contig sizing.

One-percent agarose (SeaKem LE; FMC BioProducts) gels are prepared in 1X TAE (Sambrook et al. 1989). Molten agarose is cooled to 46°C in a water bath with occasional stirring and then poured into 20 by 25-cm UV transparent trays (Life Technologies) resting on a level surface. The comb is then inserted. For each gel, 150 ml of molten agarose is used, resulting in a gel thickness of approximately 3.5 mm. The comb used should form 51 wells with the following dimensions: 2 mm wide by 1 mm thick by 3 mm deep, where thick is the dimension in the direction of DNA migration. After the gel solidifies the comb is removed, the gel is wrapped in Saran Wrap and stored at 4°C until electrophoresis. Typically this storage time period should not exceed 3 days. Gels are removed from 4°C storage and placed into electrophoresis units containing buffer at the desired electrophoresis temperature for at least 10-min before sample loading. The restriction enzyme digestion/loading dye mixture (1.75 µl) is loaded into each well. In the first well and every fifth well thereafter, 1 µl of a standard marker DNA sample is loaded. Marker DNA should be a mixture of 1 kb ladder (Life Technologies) and both Marker II and Marker III (Boehringer-Mannheim) in the

following proportions: 0.83 µl (1 µg/µl) 1 kb ladder, 3.33 µl (250 ng/µl) Marker II, 3.33 µl (250 ng/µl) Marker III, 92.51 µl TE [10 mM Tris (pH 8.0), 0.1 mM EDTA (pH 8.0)], 25 µl 6X loading dye. Immediately before electrophoresis, 20 µl of this mixture is removed to a separate tube, diluted by the addition of 17 µl of TE and 3 µl of 6X loading dye and incubated at 60°C for 5 min.

Samples are electrophoresed in Model H4 electrophoresis units (Life Technologies) at 90 V for 15 min after which time recirculation of the electrophoresis buffer (1x TAE; Sambrook et al. 1989) is initiated. Buffer is recirculated by pumping through 25 feet of small diameter Tygon tubing (Tygon LFL 6429-17) immersed in a 16-liter tank containing water maintained at a constant temperature of 14°C. Temperature regulation of the water is achieved using a refrigerated recirculator (VWR Scientific, model 1170). A tank temperature of 14°C serves to maintain a constant electrophoresis buffer temperature of 16°C. Total electrophoresis time is 8 hr.

After electrophoresis, gels are removed to plastic trays containing 400 ml of a 1:10,000 dilution of either SYBR Green (FMC BioProducts) or Vistra Green (Molecular Probes) in 1X TAE, and agitated in the dark for 30-45 min. The diluted SYBR Green I and Vistra Green solutions can be reused one time. Diluted stains are stored at 4°C in a Rubbermaid (recycle number 5) container wrapped in foil. After staining, gels are imaged using a Molecular Dynamics FluorImager SI with the following scan settings: pixel size, 200 µm; digital resolution, 16 bits; detection sensitivity, high; PMT voltage, 950 V; Filter, 530 nm. Gel images are first cropped and then converted from the proprietary 16-bit Molecular Dynamics format to 8-bit TIFF images, and transferred by ftp to Unix workstations for band calling and contig building. The Molecular Dynamics FluorImager is also used to measure the yield of DNA, prepared as described above, using protocols and Pico Green stain obtained from Molecular Dynamics.

D. Computer Analysis and Contig Construction

Following agarose gel electrophoresis, the restriction digests are imaged and the fingerprint data used to computationally determine the relationships between all of the clones. The resulting relationship matrix represents the clone-based physical map of the genome. From this matrix, a minimal set of overlapping clones, or tiled clones, representing the genome laid end to end can be selected and placed adjacent to each other in the multi-well plates.

Identification of restriction fragment bands is preferably performed interactively using an unmodified implementation of the program Image 2.0 (F. Wobus and R. Durbin, unpubl.) and subsequently Image 3.3 (D. Platt, F. Wobus, and R. Durbin, unpubl.), suitably modified to accept gel images generated as described above. Band call data are collected and used to perform contig assembly in the program FPC (C. Soderlund and I. Longden, Sanger Centre Technical Report SC-01-96, August 1996) using functions available in FPC and the program MAPSUB (Sulston et al. 1988). Image and FPC have been developed and are maintained at the Sanger Centre; documentation and user's manuals are available from the Sanger Centre website (http://www.sanger.ac.uk).

For PAC and BAC clones, first select a clone (clone 1) and compare it to all clones in the FPC database using parameters of tolerance=7, cutoff score=$10^8$. The term tolerance refers to a window size; for example, if tolerance is set at 7, then two restriction fragments occurring in different fingerprints must have relative mobilities within seven-tenths of a millimeter to be considered equivalent fragments. A decrease in tolerance decreases the window size and therefore, increases the stringency of the comparison. It is important to note that all of the calculations performed in FPC have used the relative mobilities of the restriction fragments and not the sizes of the restriction fragments.

The cutoff score is a threshold value representing the maximum allowable probability of a chance match between any two clones (the Sulston score). The smaller the Sulston score value, the lower the probability that the match has arisen by chance, and the more extensive the overlap between any two clones. Practical experience with extensive fingerprint data has led us to apply a cutoff score of $10^8$. Details describing the derivation of the scores and issues relating to the calculation of the Sulston score are presented by Sulston et al. (1988). Matches between clone 1 and other clones are displayed. Select the clone (clone 2) exhibiting the best match (i.e., the matching clone exhibiting the smallest Sulston score) to clone 1 and manually compare, using a fingerprint viewing tool provided by FPC, its fingerprint to that of clone 1 to determine the number of shared fragments. The overlap between the clones can then be drawn manually in FPC. If the clone 2 fingerprint exhibits no unique restriction fragments, bury (hide) clone 2 within clone 1. If unique fragments are observed in clone 2, repeat the entire procedure-using clone 2 for the next search against the FPC database. The best match (clone 3) is identified, and its fingerprint is compared manually against the fingerprints of clone 1 and clone 2. To incorporate clone 3 into the nascent contig, require

that the unique restriction fragments exhibited by clone 2 be present in clone 3. This constraint is imposed to ensure the internal consistency of the nascent contig and to provide additional assurance, through redundancy, that the clones represent faithfully the underlying genome. If this constraint cannot be met (a possibility that might arise because of, for example; a restriction fragment length polymorphism, (RFLP)) the clone may still be incorporated into the contig and used as a mapping reagent. As a precaution, the clone in question should be labeled with a tag in FPC so that it will not be selected for other manipulations including DNA sequencing. For PAC clones, which possess two variably sized vector-insert junction fragments, allow two unconfirmed fragments per fingerprint.

The process of consecutive searches continues until no matches better than the cutoff score can be identified and the contig cannot be extended further. An additional search, using the entire contig to query FPC, is performed to identify any remaining matching clones. If any are found they are incorporated into the contig as described above.

If fingerprints from multiple clonal isolates from each well address were generated, only one of the replicate fingerprints is incorporated into the nascent contig during contig assembly. The selection of the appropriate fingerprint, in the cases where differences are observed among the three fingerprints, is constrained to preserve the internal consistency of the contig. That is, all fragments (except for the two possible vector-insert junction fragments) of a clone occupying an internal position in the contig are verified manually by direct comparison with the fragments of flanking clones. To declare overlap between any two clones ~50% of the bands need to be identified as common. In the context of a contig larger than two clones, this parameter can, in practice, often be relaxed provided the constraint of internal consistency within the contig is met and new bands evident in a pair-wise comparison between two clones are confirmed by the next clone entering the contig. A minimal tiling path is selected from the contig such that the selected clones encompass all restriction fragments across the contig with minimal duplication of coverage.

A final physical confirmation of the contigs and their overlapping sequences can be accomplished by sequencing the ends of the restricted DNA fragments.

EXAMPLE 2

1. Construction of Genomic DNA Microarray Chips

As illustrated in Figure 3, individual genomic DNA cloned sequences from the selected area 26 of the genomic DNA or chromosome 24 are used to construct the genomic

DNA microarray chip 30. Each (bolded) genomic DNA cloned sequence 32 in the tiling path of a genome will be arrayed 36 in the appropriate area 37 on the solid surface 34, such as a standard glass microscope slide, using conventional techniques. Current arraying technology, such as that developed by Molecular Dynamics, Inc. of Sunnyvale, CA, provides for the deposition of over 9,000 DNA samples on a microscope slide. Thus, the tiling path for a large size genome is easily contained on such arrays. Genomic DNA microarray chips allow the construction of useful arrays for complete and comprehensive gene expression analysis without the need for costly, inaccurate and time-consuming characterization of all genes within the genome of interest. The ability to efficiently interrogate all of the genes of a genome of interest has not yet been realized using the conventional current chip technology and configurations currently in use for gene expression analysis.

A unique feature of employing the genomic DNA microarray chips is that the hybridization of labeled targets can be observed at more than one microscopic optical level. Using a scanning confocal fluorescence microscope, the observation of the hybridized-labeled targets is seen as a single fluorescent spot on the solid surface. When using a phase fluorescent microscope on the same genomic DNA microarray chip, individual DNA molecules can be observed in the field of each spot that have been hybridized with labeled targets. Using the high range of microscopic objectives, differences in the hybridization patterns of exons or other probed areas can be seen on the individual DNA molecules.

A. Derivatized Glass Surface Preparation.

A glass microscope slide is used as the solid support surface onto which the arrayed genomic DNA clones are placed. The glass microscope slide is prepared using the following method.

Glass microscope slides are cleaned in nitric acid overnight, then in hydrochloric acid for another 6 hours, followed by a thorough rinse in deionized water. Slides are dried in air and then incubated with 0.1% aminopropyltriethoxysilane (APS) (Aldrich) in 95% ethanol for 10 minutes at room temperature. Derivatized slides are rinsed with deionized water and air-dried.

B. DNA Preparation, Mounting for VTM and VTP

5ul of the intercalating dye YOYO (0.2um; Molecular Probes) is used to stain BAC DNA (0.05ng/ul) in TE buffer. The stained BAC DNA is pippetted (open-bore tip) onto polylysine slides and a cleaned 22X50mm coverslip is applied on the top of the slide. Force

is used to spread the solution drop of DNA on the surface of the slide. Pressure is applied until fringes appear around the edge of the coverslip. The coverslip is then separated from the slide after a few minutes. Slides are allowed to dry at room temperature and then baked at 60° C for at least 4 hours and up to overnight. DNA molecules thus stretched and fixed are observed under a fluorescence microscope and are ready for VTM and VTP experiments.

C. Fixation of Arrayed DNA Samples.

The purified genomic DNA clones are restricted, fixed and aligned in ordered square arrays that mimic the clone locations set up originally in the multi-well plates. This matched array is used for comparing the hybridization patterns with the proper clonal recognition. This array may represent a part of or the total genome.

DNA molecules are first linearized by digestion with a suitable rare cutter restriction enzyme that has a recognition site in the multiple cloning site of the vector. DNA molecules are elongated and aligned in square arrays by spotting droplets of DNA solution onto the derivatized glass surfaces, followed by air drying, using an Eppendorf micro manipulator in combination with a x-y table (interfaced to a computer) controlled by microstepper motors. A glass capillary tube (500 μm, i.d.) can be used to draw DNA samples and then spot onto derivatized glass surfaces by simple contact. Each spot will be typically 900 μm with a spot to spot variation of ±100 μm. A center to center spacing between spots of 1.5 mm is controlled by computer program settings of the micromanipulator, and x-y table combination. Alternatively, grids can be generated by using a modified commercially available laboratory automation robot equipped with a 500 μm ID stainless steel capillary pipetting tool, and a specialized workspace deck capable of holding multiple 96 well microtiter plates and up to 12 optical mapping surfaces in a vacuum chuck. Fluid droplets (5-50 pg/μl of DNA in Tris EDTA buffer) of 10-20 nl are spotted onto open glass surfaces that had been derivatized with APTES or [3-triethoxysilyl-propyl]trimethylammonium chloride (TESP), using customized robots for deposition of spots as described in Jing *et al.*, 1998.

2. Genome Mapping

As shown in Figure 4A, known sequences can be localized or anchored on a genomic DNA cloned sequence 40. The genomic DNA sequence 40 can be hybridized with known ESTs, cDNAs, PCR products or genomic fragments that are labeled with reporter molecules (e.g., fluorescent dyes). This genomic DNA cloned sequence 40 was derived from a selected area 26 of the genomic DNA or chromosome 24. In this case the labeled mRNA or DNA

sequences are referred to as the probe and the genomic DNA cloned sequence 40 is the target. In Figure 4A, the genomic location of Gene A 42, Gene B 44, and Gene C 46 is being sought using the labeled probes mentioned above.

As shown in Figure 4B, when probing the target genomic DNA cloned sequence 40 with known labeled ESTs, one exon 41a in Gene A 42 and three exons 41b in Gene B 44 can be visualized. When hybridizing the target genomic DNA cloned sequence with known mRNAs or cDNAs five exons 43a in Gene B 44 and three exons 43b in Gene C 44 can be located. Comparing these outcomes with bioinformatic based computer predictions might show that Gene A 42 should have three exons 45a where one matches up perfectly with the known EST 41a. Bioinformatic based computer prediction show that Gene B 44 should have five exons 45b which correlates with the five exons 43a that were seen with the hybridization of known mRNAs and cDNAs. The three exons 41b located with the known ESTs in Gene B 44 correlates well with these predictions also. In contrast, the predicted four exons 45c of Gene C 46 do not match completely with the hybridization pattern of the known mRNAs or cDNAs 43b and no known EST hybridization was seen at all within Gene C 46.

Expanding on the above example, solution hybridization of labeled probe DNA derived from coding or non-coding sequences can be used to localize the target within the complete array or tiling path. After hybridization of the labeled probe to one or more target clones, a scan of the array using an appropriate conventional imaging system (e.g., fluorescence detection) indicated which targets contain the probe sequence. *See* Heiskanen et al. (1994). These target clones can then be the focus of additional experiments (e.g., DNA sequencing) to further characterize the genomic sequence identified by the probe.

Alternatively the labeled probe need not be derived from the same genome as the arrayed oligonucleotide DNAs but could be from another species. If an analogous gene or homologue is present in the target genome, the clone or clones in which it is contained will be identified with this procedure. In this manner, known markers are assigned to the tiling path to anchor the contigs to an existing map with new markers concomitantly assigned to a map position. Hybridization of the probe to immobilized DNA is detected as a specific signal associated with a clone or overlapping clones. This feature will be advantageous in the application of data from model organism genomes to commercially important genomes such as that of human or food crops and animals.

The foregoing method of making a genomic DNA microarray chip is preferably modified by placing an entire ordered set of genomic clones on the array, in order to increase the resolution of the map position obtained following hybridization of the probe. Specifically, the overlap of clones that are identified as containing the target can be used to define precisely the map position to the minimal region in common between all of these clones.

EXAMPLE 3

1. Using the Genomic DNA Microarray Chips for Transcriptional Profiling or Gene Expression Analysis

Conventional gene chip technology is directed towards detecting changes in expression levels of specific genes in mRNA populations derived from different tissues, developmental stages or from different environmentally challenged organisms. One of the severe limitations of conventional array technology is that the probe population is derived solely from known genes. The definitions of probe and target stated previously apply here. By contrast, the genomic DNA microarray chip teaches that the whole or a partial genome of an organism can be used as the probe in the form of arrayed, ordered, tiled genomic DNA clones. One important advantage of these genomic DNA arrays over currently used gene chips is that all or part of the coding segments of the genome is represented without prior knowledge of their existence.

By substituting a genomic DNA microarray chip for a conventional gene chip in a typical gene expression analysis or transcriptional profiling procedure, previously unknown genes can be identified, since they will hybridize to target molecules in a mRNA population. The probe and target nomenclature as it is applied to conventional gene chip technology constitutes somewhat of a departure from conventional thinking in that the newly discovered gene probe was previously unknown. Rather, the labeled molecule that detects the gene will be acting as a probe. However, since the genomic DNA clones that are immobilized on the array have been characterized as to their location in a particular genome, the standard chip nomenclature will be applied and the immobilized DNA will be referred to herein as the probe.

2. Direct Visualization of Individual Genes and Exons

Fluorescent signal saturation will obscure the differential expression of genes within a clone by saturating the reporter signal observed. Fluorescent signal saturation on the

genomic DNA microarray chip can occur for several reasons. For example, a high gene density in the genome of small model organisms will often yield multiple genes within any one large-insert clone and their combined signals will cause signal saturation. Likewise, when two neighboring genes are expressed at high levels, their combined signals may obscure the detection of differential expression of the two genes.

As illustrated in Figure 5, one approach to overcoming the problem of signal saturation is to use a genomic DNA microarray chip 50 containing a highly redundant set of partially overlapping clones as described in Example 1. Because the arrayed clones have a large staggered overlap (the overlapping region shown between the solid lines 57 and 58 in Figure 5), probes containing the differentially expressed gene in target A 53 and target B 55, but not the constitutively expressed gene or genes, will be present. In this manner, the differential expression will not be obscured in all clones.

As a further extension of the use of arrayed genomic clones for gene expression profiling, direct visualization of elongated, linearized large-insert probe molecules would obviate the problems described above. A process similar to conventional optical mapping (Schwartz, 1993; Samad, 1995; Schwartz, 1995; Jing, 1998) or Fiber-FISH (Heiskanen, 1995; Horelli-Kuitunen, 1999) can be used where linearized clones can be arrayed on an appropriate substrate (e.g., glass) under conditions that favor elongation of the immobilized DNA (Heiskanen, 1994). Those methods have shown that a solution containing DNA can be spotted onto glass in such a manner that labeled probes can be hybridized to the immobilized DNA.

In Figure 6, a 2.7 kb C21orf3 cDNA 62 and an 956 bp EST 7 64 is hybridized to a PAC 92C23 DNA sequence 60 (L Peltonin et al., NPHI-Helsinki). Individual exons can be observed when the cDNA 62 and the EST 64 are detected via their fluorescent reporter molecules. Using readily available commercial microscopes, images of the elongated DNA molecules can be captured.

Optical mapping methods allow the deposition of thousands of copies of a DNA clone to be examined in each spot. Hundreds of spots have been placed on a single glass slide, with even higher-density achievable. An imaging system can be used to capture an image of elongated probes having distinct regions that hybridize to the target. By using different labels for each of the target populations, differentially expressed sequences can be observed in the presence of constitutively expressed neighboring sequences. More specifically,

individual genes contained within each of the immobilized clones will be directly visualized. Furthermore, the individual exons of each gene will be resolved, allowing the detection of differential or alternative splicing. Conventional DNA chips cannot detect this important component of gene regulation, unless all exons of a gene are individually represented in the probe array. Once again, this would require that the complete gene structure of every gene on a conventional chip be known in advance.

Direct visualization of the individual genes and exons contained within a genomic clone or clones is much more accurate than existing computational tools that attempt to predict gene and exon locations on the basis of sequence composition and common characteristics. In this manner, individual exons hybridizing to reporter molecules can be counted rather than simply measuring the combined signal from all DNA molecules in the spot. This will allow more precise measurements as this converts the signal to a digital rather than an analog format.

DNA mounting and array construction have been described in the previous sections. The following outlines methods for denaturing of the DNA, probe labeling, hybridization of probe to target DNA and microscopy/imaging analysis for visual transcriptional profiling (VTP) and visual transcription mapping (VTM).

A. Probe Labeling: VTM

Amounts of 100-200 ng of each cDNA (EST) are directly labeled with fluor-12-dUTP (Stratagene) by use of random primer labeling (Stratagene; Prime-It Fluor). Large insert clones, including cDNAs, are labeled with fluor-12-dUTP by nick translation according to standard protocols (Boehringer Mannheim, Nick Translation Mix).

B. Probe Labeling: VTP

Purified mRNA is isolated using a commercial kit (*e.g.* FastTrack mRNA isolation kit) according to the manufacturer's protocol. Purified mRNA from target tissues is used to prepare fluorescently labeled cDNA for hybridization to the microarrays. Cy3-dUTP or Cy5-dUTP (Amersham) is incorporated during reverse transcription of 1.25 μg of polyadenylated [poly(A)+] RNA, primed by a dT(16) oligomer. This mixture is heated to 70°C for 10 min, and then transferred to ice. A premixed solution, consisting of 200 U of Superscript II (Gibco), buffer, deoxyribonucleoside triphosphates, and fluorescent nucleotides, is added to the RNA. Nucleotides are used at these final concentrations: 500 μM for dATP, dCTP, and dGTP and 200 μM for dTTP. Cy3-dUTP and Cy5-dUTP are used at a final concentration of

100 μM. The reaction is then incubated at 42°C for 2 hours. Unincorporated fluorescent nucleotides are removed by first diluting the reaction mixture with of 470 μl of 10 mM Tris-HCl (pH 8.0)/1 mM EDTA, and then subsequently concentrating the mix to ~5 μl, using Centricon-30 microconcentrators (Amicon).

C. Hybridization of probe to target

Purified, labeled DNA is resuspended in 11 μl of 3.5X SSC, 0.3% SDS, and 10 μg of polydeoxyadenylic acid. Blocking DNA (e.g., 20 μg of human CoT1 DNA (Gibco-BRL) for microarrays with human DNA clones) also should be included in this mixture. Slides with genomic DNA clones are denatured for 2 minutes in 70% formamide, 0.6XSSC, pH 7.0 at 72°C then put through an ice cold ethanol series (70%, 90%, 100%) for 2 minutes each and air dried. A 25 μl solution containing 20 ng of labeled probe, 30% Formamide, 1XSSC and 10% dextran Sulfate is denatured at 75°C for 5 minutes and added to the above-prepared slide. A coverslip is placed on the top of the slide and slide is sealed with rubber cement. Hybridization is carried out at 37°C for 24 hours in a humidify chamber. After the hybridization, the slide is washed with 50% formamide and 2XSSC 3X5 minutes each at at 37°C.

D. Microscopy and Imaging

Automatic imaging workstations are built around a Zeiss 135 (or equivalent) inverted microscopes equipped for epifluorescence, with 100X Zeiss plan-neofluor oil immersion objectives, numerical aperture 1.3, and multiband-pass filter pack (suitable for fluorescent labels and DNA counterstain). Preferred microscopes are equipped with a Dage SIT68GL low light-level video camera for acquiring focus, and a Princeton Instruments cooled charge-coupled device digital camera (1,316 X 1,032 pixels, KAF 1400 chip, 12-bit digitization) for high-resolution imaging and photometry. A Ludl Electronics x-y microscope stage with 0.1-μm resolution is used for translation.

DNA molecules are imaged using a software package that integrates all of the workstation functions such as movement of the microscope stage, focus, and image collection. Control of light path actuators, video auto-focus, and sample translation (x-y stage) is accomplished by a Ludl Electronics MAC 2000 interface bus with the following modules installed: PSSYST 200, MCMSE 500, MDMSP 503, AFCMS 801, FWSC 800, and RS232INT 400. The Ludl MAC 2000 is interfaced via RS232 serial connection to a Sun Microsystems SPARC 20 dual-processor computer workstation. The Princeton Instruments

charge-coupled device camera also is interfaced, via a Pentium-based microcomputer controller and distributed network, to a Sun workstation. Software for control of the above peripherals is written in the C programming language.

Digital images are acquired by the workstation and stored on hard-disk arrays for image processing and extraction of transcript mapping data. The system runs on a network of Sun workstations with a networked file system.

Images are analyzed by locating specific hybridization signals from the labeled molecules on the elongated DNA molecules. The positions of the specific signals are measured from each end of the elongated DNA molecule. The average measurement from multiple DNA molecules showing hybridization is used to position the point of signal hybridization.

3. Visual Transcriptional Profiling (VTP)

Visual Transcription Profiling (VTP) is used for measuring differential expression of genes in two mRNA samples isolated from different developmental, growth or stress states. Each population of mRNA is labeled with different reporter molecules. Figure 7 illustrates the hybridization of differentially labeled mRNA populations to a genomic DNA microarray chip 70 from the selected area 26 of the genomic DNA or chromosome 24. Different colored labels can be assigned for each plant cell mRNA populations to be compared, (e.g., Target A a normal mRNA population 72 or Target B mutant phenotype mRNA population 74). These two populations of labeled targets are hybridized to the DNA on the same chip or different chips. Changes in the gene expression between these two mRNA populations can be detected with colored fluorescent reporter labels that are attached to the mRNA molecules. Differential expression is observed by evaluating the hybridization signals obtained for the unique hybridization of the target molecules to an arrayed DNA or variations in the co-localization of the reporter molecules when compared to other mRNA targets. This evaluation of signals performed using the same techniques as employed with standard DNA microarrays, *See* Duggan et al. (1999).

A unique feature of employing the genomic DNA microarray chips is that the hybridization of labeled targets can be observed at more than one microscopic optical level. Using a scanning confocal fluorescence microscope, the observation of the hybridized-labeled targets is seen as a single fluorescent spot on the solid surface. When using a phase fluorescent microscope on the same genomic DNA microarray chip, individual DNA

molecules can be observed in the field of each spot that have been hybridized with labeled targets. Using the high range of microscopic objectives, differences in the hybridization patterns of exons or other probed areas can be seen on the individual DNA molecules.

In gene expression analysis, differences in the hybridization patterns observed between target A 72 and target B 74 in mRNA would indicate that a gene or genes contained in the positive probe DNA potentially exhibit differential expression in the two targets. This is a significant advance over existing chip technologies as it allows the identification of genes which have not been previously cloned or characterized as well as their immediate association with diseases or particular traits (phenotypes).

4. Visual Transcript Mapping (VTM)

Besides representing a powerful tool for the analysis of gene expression, Visual Transcript Mapping can use the genomic DNA microarray chips to provide a novel approach and technique for initial gene discovery and characterization. Previously undiscovered genes can be found using the genomic DNA microarray chips as the probe for hybridization against total labeled mRNAs.

After a genomic DNA microarray chip has been used to find and pinpoint a potentially interesting gene or region of the genome, more focused DNA sequencing work can be done to further characterize the gene that has been found within a specific genomic interval. For an investigator interested in such a gene or genomic region, this improved technology would provide considerable savings in terms of both cost and time.

Although the present invention and its advantages have been described in detail, it should be understood that various changes, substitutions, and alterations can be made to the described method without departing from the spirit and scope of the invention as defined by the appended claims.

## REFERENCES

Brown, P.O. and Botstein, D. (1999). Exploring the new world of the genome with DNA microarrays. *Nature Genet. Supp.* 21, 33-37.

Cheung, V.G., Morley, M., Aguilar, F., Massimi, A., Kucherlapati, R., Childs, G. (1999) Making and Reading Microarrays. *Nature Genet. Supp.* 21, 15-19.

Duggan, D.J., Bittner, M. Chen, Y. Meltzer, P. and Trent, J.M. (1999). Expression profiling using cDNA microarrays. *Nature Genet. Supp.*21, 10-14.

Fodor SP, Read JL, Pirrung MC, Stryer L, Lu AT, Solas D. (1991). Light-directed, spatially addressable parallel chemical synthesis. *Science* 251, 767-773.

Heiskanen, M., Karhu, R., Hellsten, E., Peltonen, L., Kallioniemi, O.P. and Palotie, A. (1994). High Resolution Mapping Using Fluorescence *in Situ* Hybridization to Extended DNA Fibers Prepared from Agarose-Embedded Cells. *BioTechniques* 17, 928-933.

Heiskanen, M., Hellsten, E., Kallioniemi, O.P., Makela, T.P., Alitalo, K., Peltonen, L., Palotie, A. (1995). Visual mapping by fiber-FISH. *Genomics* 30, 31-36.

Horelli-Kuitunen ,N., Aaltonen, J., Yaspo, M.L., Eeva, M., Wessman, M., Peltonen, L., Palotie, A. (1999). Mapping ESTs by fiber-FISH. *Genome Res.* 9,62-71.

Jing, J., Reed, J., Huang, J., Hu, X., Clarke, V., Edington, J., Housman, D., Anantharaman, T.S., Huff, E.J., Mishra, B., Porter, B., Shenker, A., Wolfson, E., Hiort, C., Kantor,R., Aston, C., Schwartz, D.C. (1998). Automated high resolution optical mapping using arrayed, fluid-fixed DNA molecules. *Proc. Natl. Acad. Sci. USA* 95, 8046-8051.

Lipshutz, R., Fodor, S.P.A., Gingeras, T.R., Lockhart, D.J. (1999) High Density Synthetic Oligonucleotide Arrays. *Nature Genet. Supp.* 21, 20-24.

Marra, M., Kucaba, T., Dietrich, N., Green, E., Brownstein, B., Wilson, R., McDonald, K., Hillier, L., McPherson, J. and Waterston, R. (1997). Agarose gel-based high throughput fingerprint analysis of large insert clones: Contig construction and selection of clones for large scale DNA sequencing., *Genome Research* 7, 1072-1084.

Samad, A., Huff, E.F., Cai, W., Schwartz, D.C. (1995). Optical mapping: a novel, single-molecule approach to genomic analysis. *Genome Res.* 5, 1-4

Schena, M., Shalon, D., Davis, R.W., Brown, P.O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270, 467-470.

Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98, 503-517.

Schwartz, D.C., Li, X., Hernandez, L.I., Ramnarain, S.P., Huff, E.J., Wang, Y.K. (1993). Ordered restriction maps of Saccharomyces cerevisiae chromosomes constructed by optical mapping. *Science* 262, 110-114

Schwartz, D.C., Samad, A. (1995). Optical mapping approaches to molecular genomics. *Curr. Opin. Biotechnol.* 8,70-74.

All patents and publications mentioned in this specification are indicative of the level of skill of those of knowledge in the art to which the invention pertains. All patents and

publications referred to in this application are incorporated herein by reference to the same extent as if each was specifically indicated as being incorporated by reference and to the extent that they provide materials and methods not specifically shown.

CLAIMS

1. An array of oligonucleotides immobilized on a solid support, said array comprising a group of DNA clones having minimally overlapping sequences and together making up an ordered tiling path representing at least part of an entire genome, at least one of said oligonucleotides being of unknown DNA sequence.

2. The immobilized array of claim 1 wherein said group of DNA clones represents an entire genome.

3. A method of determining the genomic location of a labeled DNA sequence comprising hybridizing a labeled DNA sequence chosen from the group consisting of an EST, cDNA, PCR product or genomic fragment to at least one DNA clone on the immobilized array of claim 1.

4. In a method of detecting a change in gene expression including hybridizing mRNA to at least one immobilized DNA sequence, an improvement comprising substituting the immobilized oligonucleotide array of claim 1 for an immobilized array of genes of known DNA sequence.

5. In a method of detecting differential gene expression in at least two populations of mRNA, the method including hybridizing mRNA to at least one immobilized DNA sequence, an improvement comprising substituting the immobilized oligonucleotide array of claim 1 for an immobilized array of genes of known DNA sequence.

6. A method of preparing the array of claim 1, comprising:

preparing a plurality of genomic DNA fragments;

cloning each of said genomic DNA fragments;

isolating each cloned DNA fragment;

identifying an ordered tiling path of said DNA fragments; and

immobilizing the DNA fragments in a sequential alignment representing their position in the ordered tiling path.

7. The method of claim 6, wherein the genomic DNA fragments are prepared by randomly shearing genomic DNA.

8. The method of claim 6, wherein the genomic DNA fragments are prepared by partially cleaving the genomic DNA using restriction enzymes.

9. The method of claim 6, wherein the genomic DNA fragments are cloned using a large-insert vector.

10. The method of claim 6, wherein the genomic DNA fragments have an overlapping oligonucleotide sequence.

11. The method of claim 10, wherein the ordered tiling path is identified by comparing the restriction pattern fingerprints of said DNA fragments.

12. The method of claim 10, wherein the ordered tiling path is identified by sequencing at least one end of each DNA fragment and aligning said DNA fragments according to said overlapping oligonucleotide sequence.

13. A genomic microchip comprising an immobilized array of DNA clones, said DNA clones including the oligonucleotide sequence of substantially an entire genome, wherein said DNA clones are immobilized in an order reflecting the sequence of said DNA clones in the genome.

14. The genomic microchip of claim 13, wherein said DNA clones are prepared using a large-insert vector.

15. The genomic microchip of claim 13, wherein the vector is a bacterial artificial chromosome.

16. The genomic microchip of claim 13, wherein said DNA clones have a partial overlapping oligonucleotide sequence.

17. A kit for determining the genomic location of known oligonucleotide sequences comprising:

a genomic chip comprising an immobilized ordered array of DNA clones derived from a plurality of DNA fragments of a genomic DNA, wherein said ordered array reflects the sequence of said DNA fragments in said genomic DNA; and

a labeled sample of a known oligonucleotide sequence wherein said known oligonucleotide sequence will hybridize to at least one of said DNA clones.

18. The kit of claim 17 wherein said known oligonucleotide sequence in an EST, cDNA, PCR product or genomic fragment.

19. A method for detecting a change in gene expression in at least two populations of mRNA comprising:

preparing a genomic chip comprising an immobilized ordered array of DNA clones derived from a plurality of DNA fragments of a genomic DNA, wherein said ordered array reflects the sequence of said DNA fragments in said genomic DNA;

preparing a first cDNA from a first population of mRNA;

preparing a second cDNA from a second population of mRNA, wherein said second cDNA is labeled;

hybridizing said first cDNA to the genomic chip;

hybridizing said second labeled cDNA to the genomic chip; and

detecting the second cDNA.

20. The method of claim 19, wherein the plurality of DNA fragments contain overlapping oligonucleotide sequences.

21. The method of claim 19, wherein the plurality of DNA fragments represents the entire genome.

22. The method of claim 19, wherein said labeled cDNA has a fluorescent label .

23. A clone array comprising a multiwelled plate having one DNA clone in each well of said plate, wherein each DNA clone is selected from a plurality of cloned genomic DNA fragments and has an overlapping nucleotide sequence with the DNA clone in at least one adjacent well, wherein said selected array of genomic DNA fragments include the entire genomic oligonucleotide sequence.

24. A method of preparing a genomic microchip comprising:

randomly shearing genomic DNA to produce a plurality of genomic DNA fragments;

cloning each of said genomic DNA fragments using a large-insert vector;

isolating each cloned DNA fragment;

preparing restriction pattern fingerprints of said DNA fragments;

determining an ordered tiling path of the DNA fragments wherein said ordered tiling path reflects the sequence of DNA fragments in said genomic DNA;

selecting a plurality of said DNA clones wherein each clone contains an overlapping nucleotide sequence with at least one other selected DNA clone and the majority of selected DNA clones contain an overlapping sequence with two other selected DNA clones; and

immobilizing the selected DNA clones in a sequential alignment representing the position of the selected DNA fragments in the ordered tiling path.

25. A genomic microchip comprising:

a plurality of purified genomic DNA fragments, wherein said genomic DNA fragments contain the entire oligonucleotide sequence of a target chromosome or genome; and

a solid glass surface to which said genomic DNA fragments are attached in an alignment reflecting the oligonucleotide sequence of the target genome.

26. A method for detecting a change in gene expression in at least two populations of mRNA comprising:

preparing a genomic microchip comprising an immobilized ordered array of DNA fragments derived from a genomic DNA, wherein said ordered array reflects the sequence of said DNA fragments in said genomic DNA;

preparing a first population of mRNA;

labeling said first population with a first label;

preparing a second population of mRNA, wherein said second population is labeled with a second label;

hybridizing said first population to the genomic chip;

hybridizing said second population to the genomic chip; and

detecting the differential labeling of the two populations.

27. A method of determining the genomic location of a labeled DNA sequence comprising:

preparing an ordered array of DNA clones, said DNA clones including the oligonucleotide sequence of substantially an entire genome, wherein said ordered array is aligned in an order reflecting the oligonucleotide sequence of said genomic DNA;

hybridizing said DNA clones with a labeled DNA sequence chosen from the group consisting of an EST, cDNA, PCR product or genomic fragment; and

identifying the DNA clone hybridized with the labeled DNA sequence.

28. The method of claim 27, further comprising the step of visualizing an expressed sequence of DNA on the hybridized DNA clone.

29. The method of claim 27, further comprising the step of visualizing a non-expresses DNA sequence on the hybridized clone.

30. A method of determining the genomic location of a designated DNA sequence comprising:

immobilizing an array of DNA clones on a solid support, said DNA clones having minimally overlapping sequences and together making up an ordered tiling path representing at least part of an entire genome, at least one of said DNA clones being of unknown DNA sequence;

hybridizing a labeled DNA sequence to said array of DNA clones wherein said labeled DNA sequence will hybridize to the designated DNA sequence; and

visualizing the labeled DNA sequence on the clone with an imaging system.

31. The method of claim 30, wherein the imaging system identifies a hybridized clone and the specific location of the hybridization on the hybridized clone.

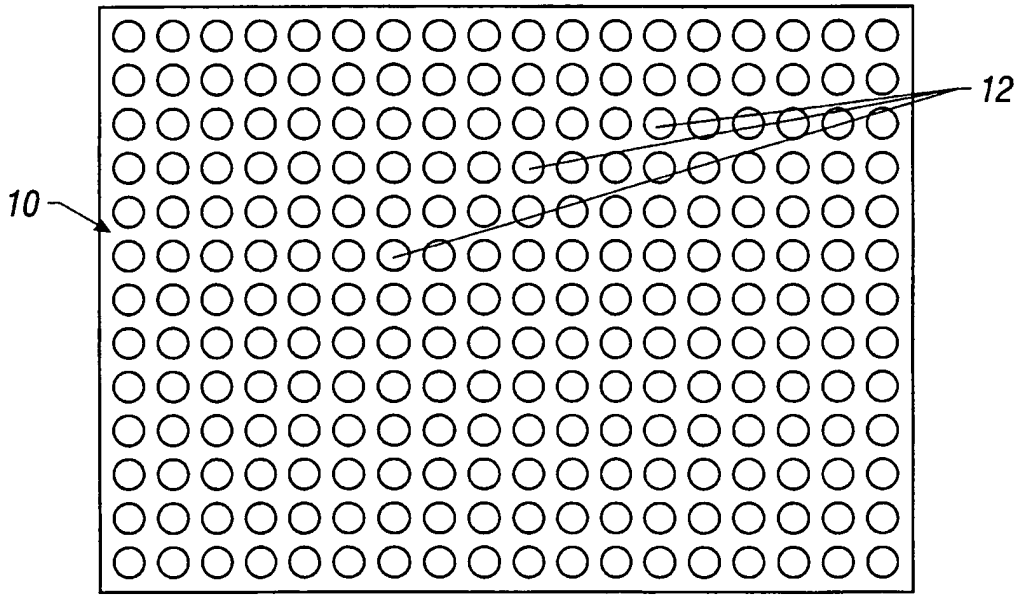


FIG. 1

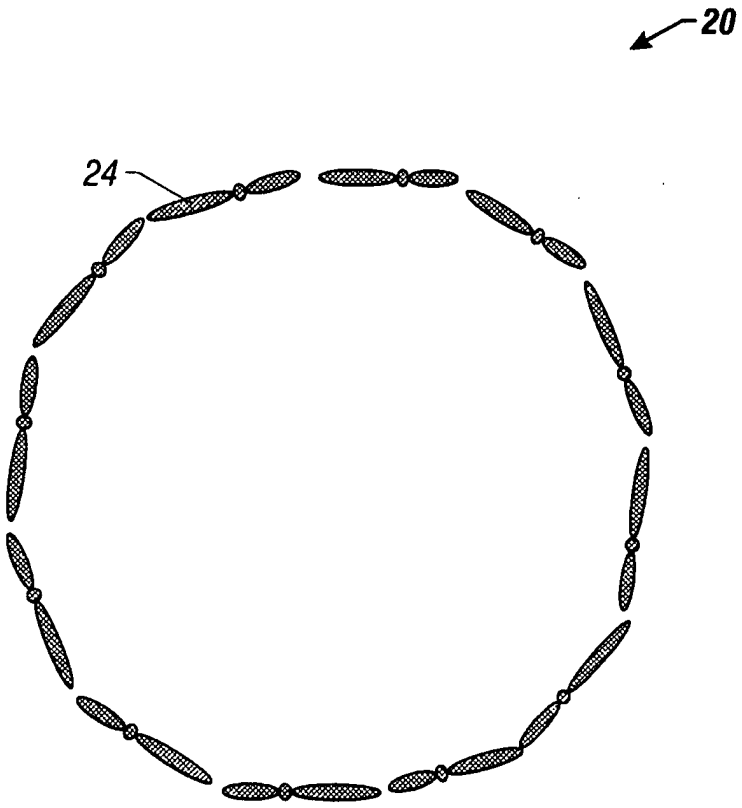


FIG. 2A

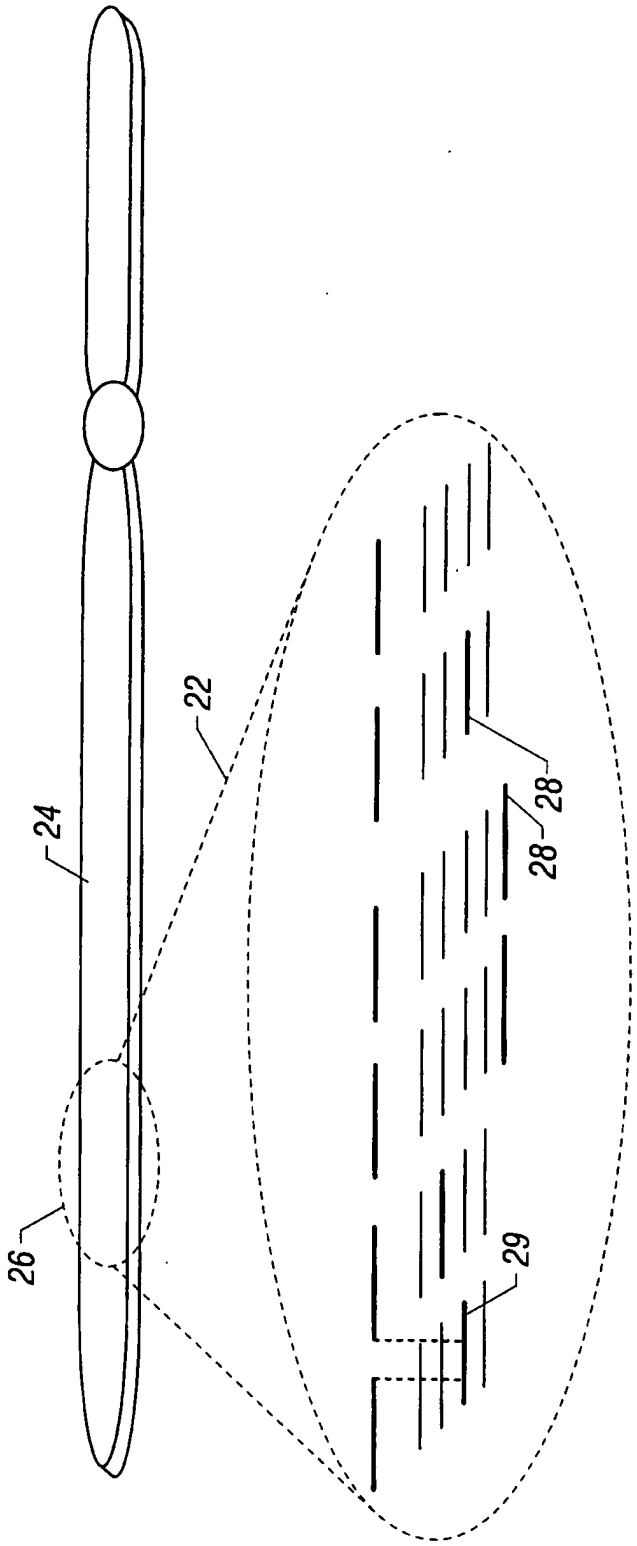


FIG. 2B

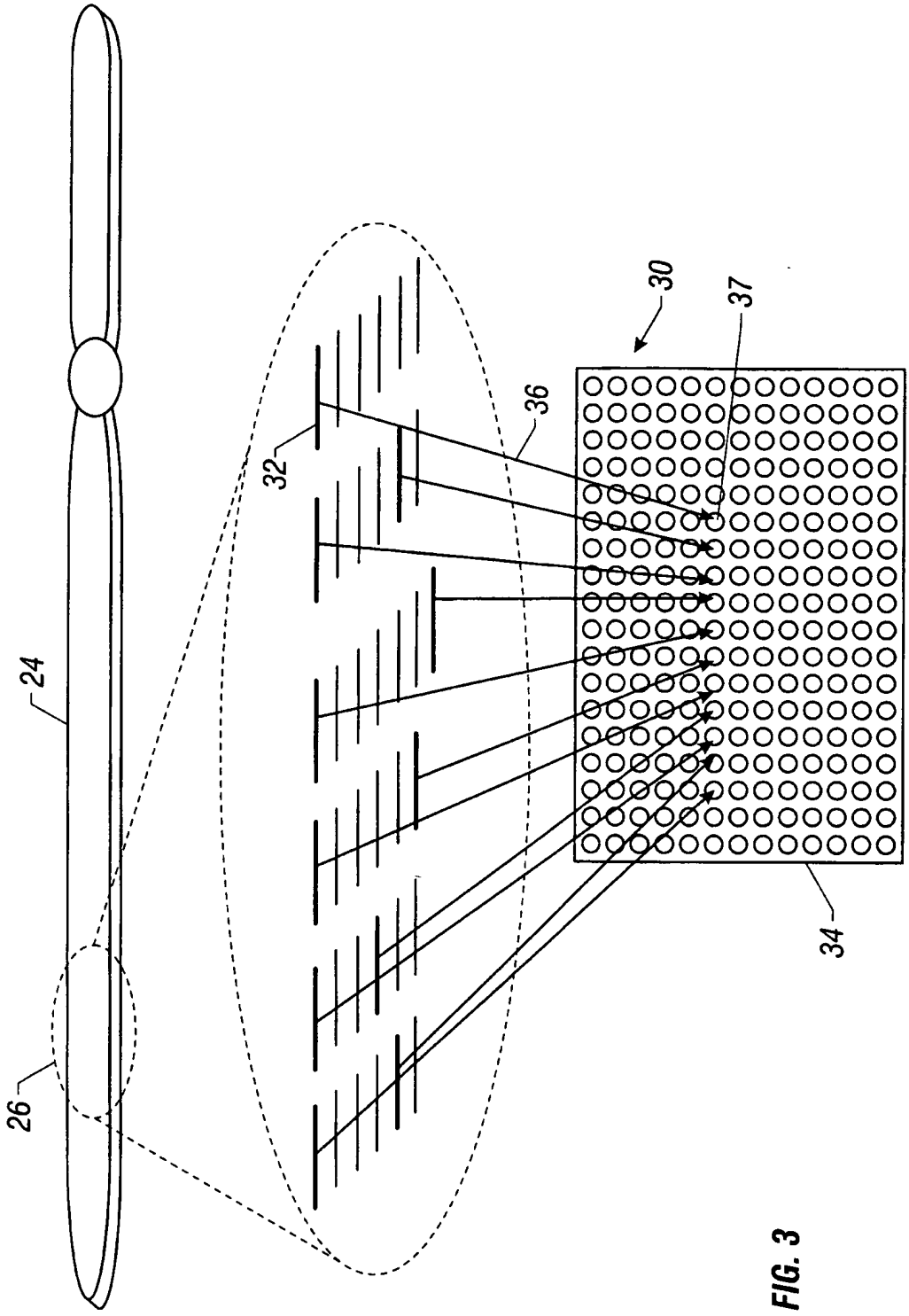


FIG. 3

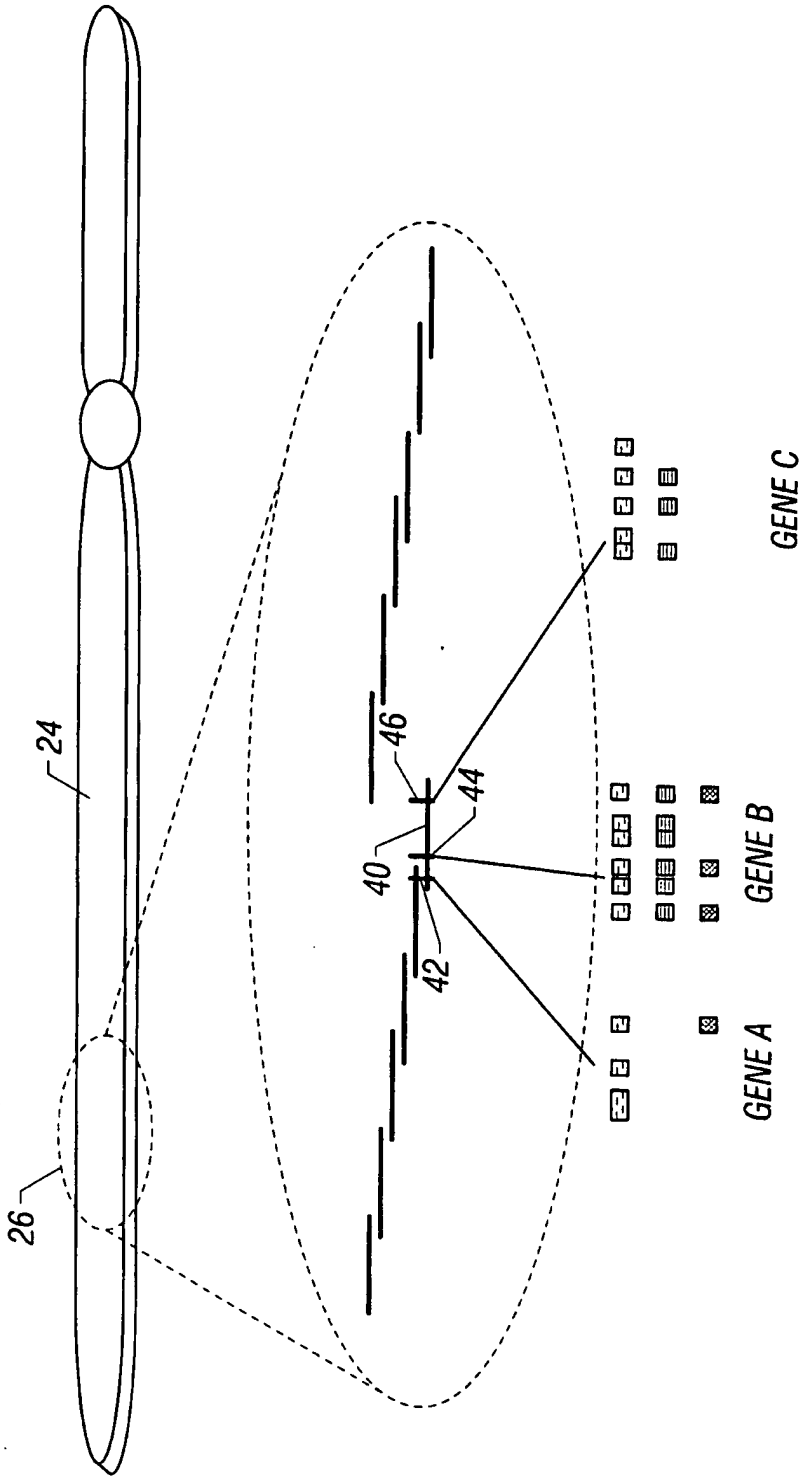


FIG. 4A

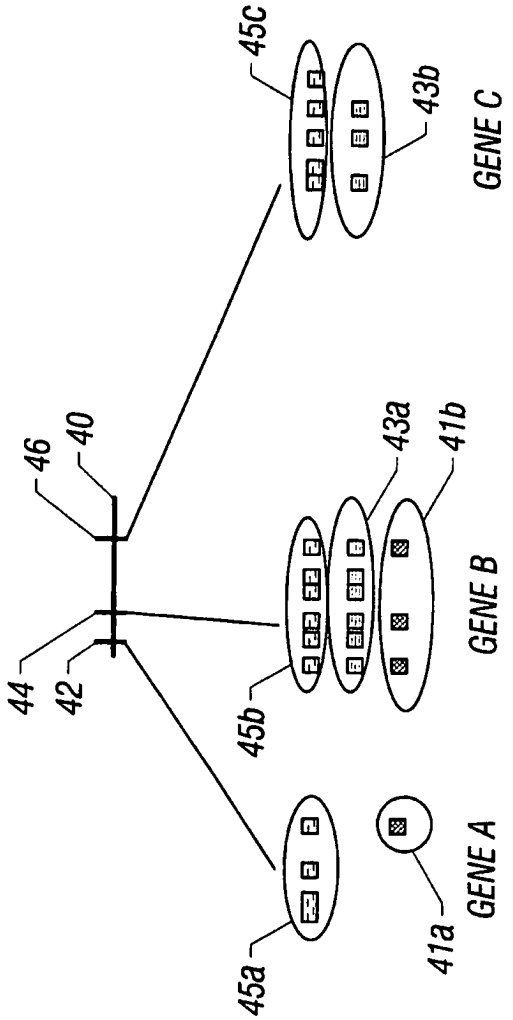


FIG. 4B

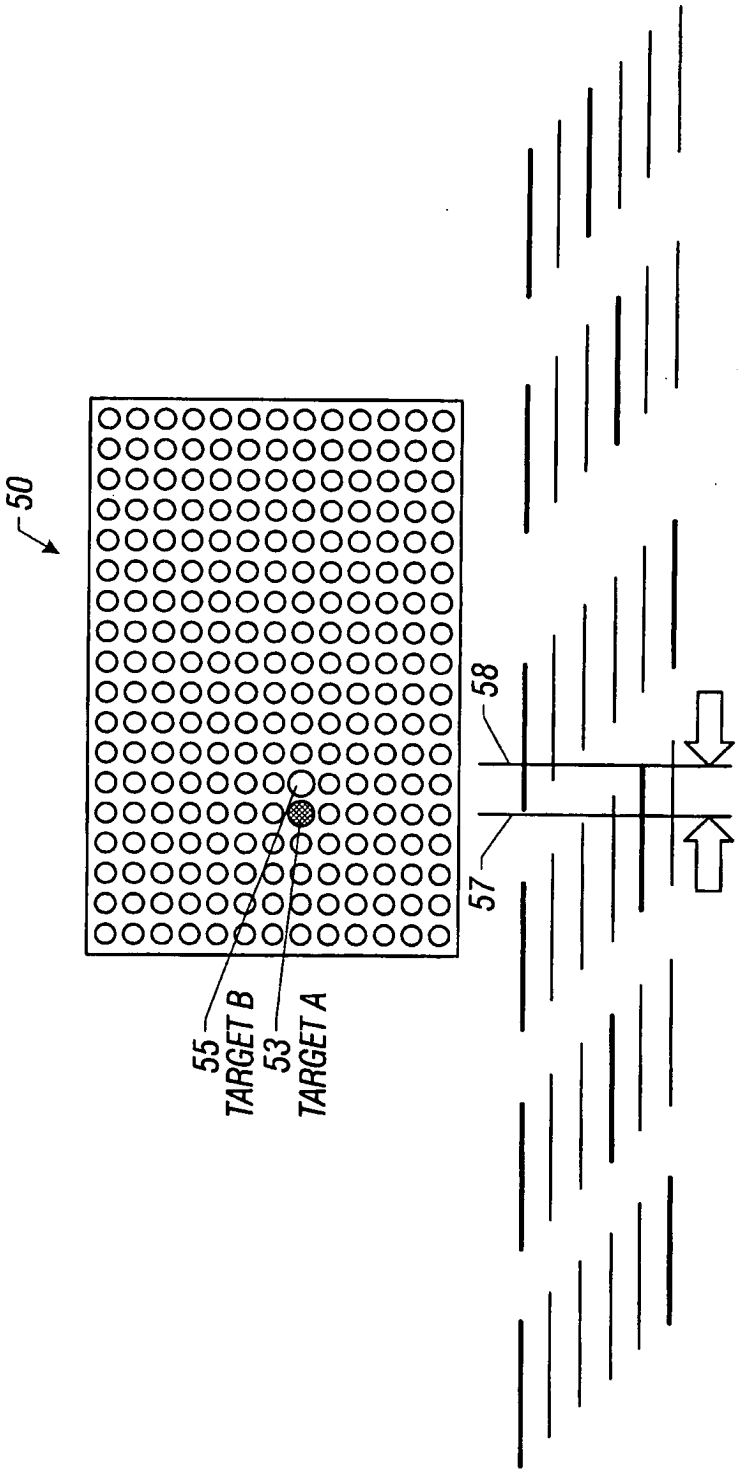


FIG. 5

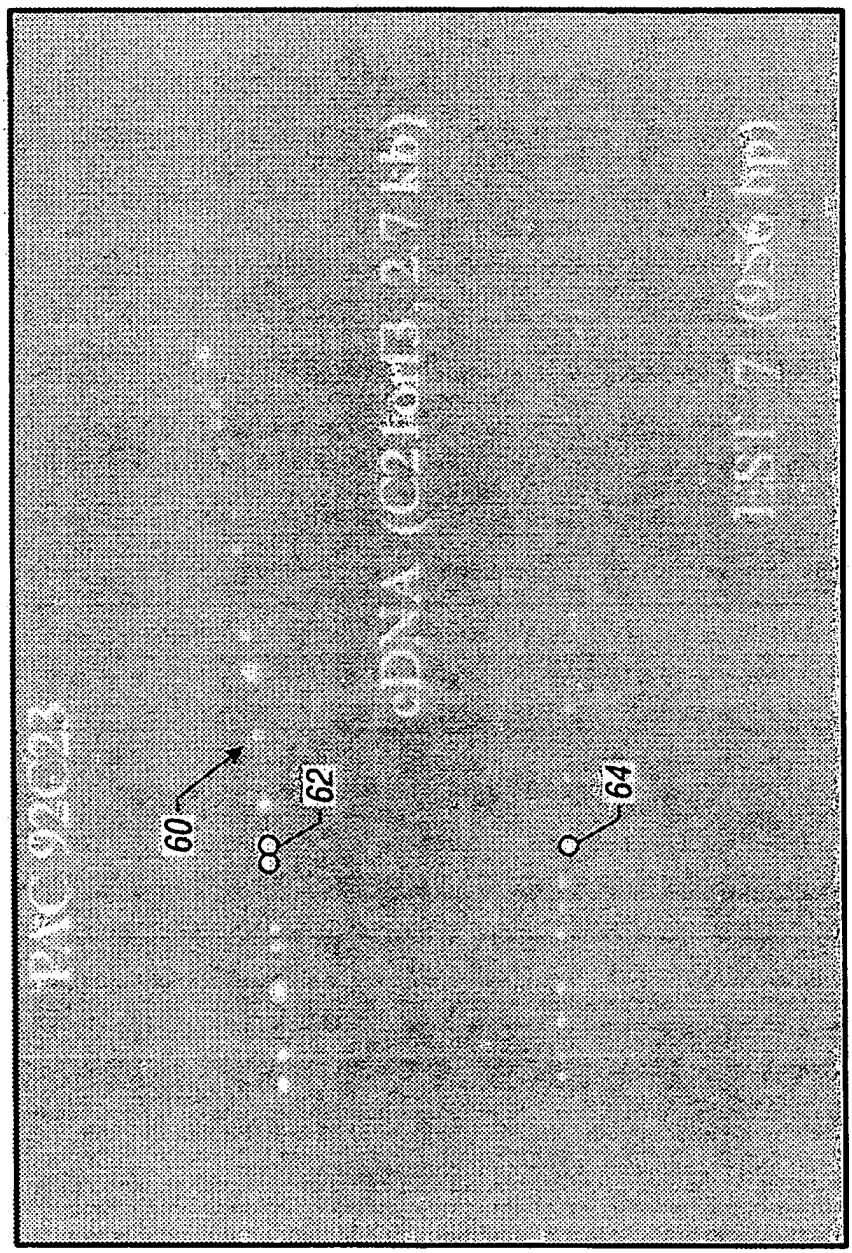


FIG. 6

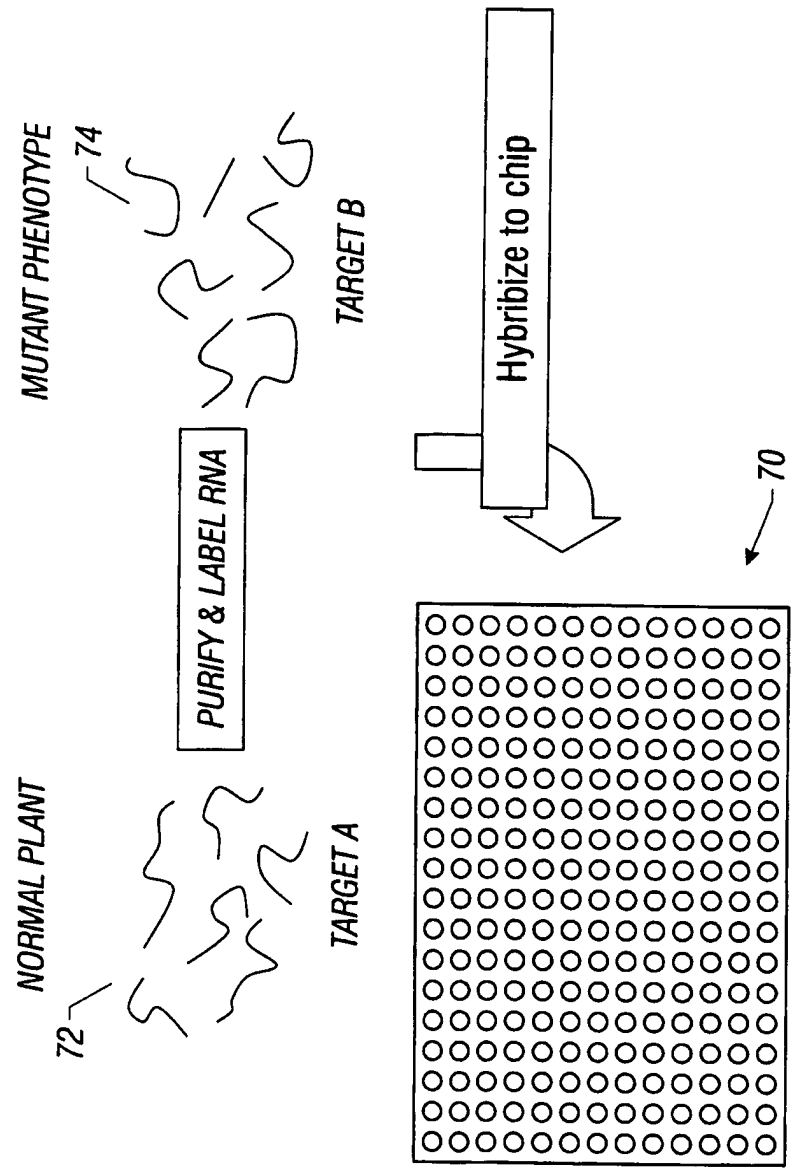
NORMAL PLANT
MUTANT PHENOTYPE
72
74
PURIFY & LABEL RNA
TARGET A
TARGET B
Hybribize to chip
70

FIG. 7

# INTERNATIONAL SEARCH REPORT

| Internationa. plication No. |
|---|
| PCT/US00/06342 |

**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(7) : C12Q 1/68
US CL : 435/6

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 435/6

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
BIOSIS CAPLUS MEDLINE SCISEARCH

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X,E | US 6,040,138 A (LOCKHART et al) 21 March 2000 (21.03.2000), column 2, line 35 to column 8, line 9. | 1-2, 13-19, 23, 27 |
| Y,E | US 6,054,270 A (SOUTHERN et al) 25 April, 2000 (25.04.2000), see entire document. | 1-31 |

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 22 May 2000 (22.05.2000) | 27 JUN 2000 |
| Name and mailing address of the ISA/US<br>Commissioner of Patents and Trademarks<br>Box PCT<br>Washington, D.C. 20231<br>Facsimile No. (703)305-3230 | Authorized officer<br>Michael Woodward<br>Telephone No. (703) 308-0196 |

Form PCT/ISA/210 (second sheet) (July 1998)